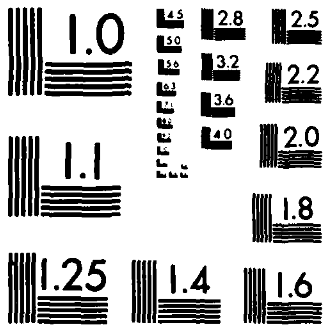
AD-A142 662 SPECULATIONS ON THE FUTURE OF TEST DESIGN(U) 1/1
EDUCATIONAL TESTING SERVICE PRINCETON NJ I I BEJAR
APR 84 ETS-RR-84-13-ONR N00014-83-C-0761

UNCLASSIFIED F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

SPECULATIONS ON THE FUTURE OF TEST DESIGN

Isaac I. Bejar

Prepared under the contract number N00014-83-C-0761
NR 150-531 with the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Educational Testing Service
Princeton, New Jersey
June 1984

SECURITY CLASSIFICATION OF THIS PAGE [illegible]

REPORT DOCUMENTATION PAGE	
1. REPORT NUMBER	AD-A142662
4. TITLE (and Subtitle) [illegible]	[illegible]
	6. PERFORMING ORG. REPORT NUMBER [illegible]
7. AUTHOR(s) [illegible]	8. CONTRACT OR GRANT NUMBER(s) [illegible]
9. PERFORMING ORGANIZATION NAME AND ADDRESS [illegible]	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS [illegible]	12. REPORT DATE [illegible]
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release [illegible]
whole or in part [illegible]
Government [illegible]

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

In S. Embretson (Ed.), [illegible]
education, and [illegible]

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Test [illegible]

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The paper [illegible]
cognitive [illegible], and [illegible]
though adaptive testing [illegible]
achievement, it will be [illegible]
psychological [illegible]
art. Such an effort [illegible]
as well as [illegible]

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

psychometric parameters, specially difficulty. This approach opens the possibility of generating items with better control of their psychometric characteristics and ultimately the development of computer-based tests that are solidly anchored in psychological theory.

Abstract

The paper discusses the potential benefits of integrating technology, cognitive science, and psychometric theory. It is argued that even though adaptive testing, as currently implemented, is an important achievement, it will be necessary to pay close attention to the psychological foundation of tests to continue advancing the state of the art. Such an effort requires construct validation in the broadest sense, as well as focusing on items and why they differ with respect to psychometric parameters, specially difficulty. This approach opens the possibility of generating items with better control of their psychometric characteristics and ultimately the development of computer-based tests that are solidly anchored in psychological theory.

Speculations on the Future of Test Design

Isaac I. Bejar

Introduction

I am grateful for the opportunity to write the final chapter for a book concerned with the improvement of test design. I do not envy for one moment the task of the contributors to this volume, for theirs is a difficult responsibility. By contrast, my task is to speculate on the future of test design, not so difficult a task when, as in this case, the contributors have provided such stimulating descriptions of their research programs.

The chapter is divided in two major sections. In the first section I identify three areas of test design that are bound to be significantly influenced by the increasing availability of technology. These three areas are computer-assisted test assembly, computer-assisted test administration, and computer-assisted test generation. All three will be significantly affected by the sheer presence of technology and thus there is the danger that they may be affected only in superficial ways. Contributions such as the ones presented in this volume will be largely responsible for effecting the hoped-for fundamental change. The second section argues that a fundamental change is more likely to come about by an integration of cognitive psychology and psychometric theory.

Technology and Test Design

Future test designers will have at their disposal the ever-growing fruits of the information revolution. The evidence for this revolution is everywhere, but most significantly it is evidenced by the increasing

presence of microcomputers at school, at home, and at work. For test designers, the increasing availability of technology is a mixed blessing. Although such growth creates the opportunity to develop better tests or administer them more efficiently, it also creates a pressure to computerize tests and use technology superficially. Three areas of test design that are vulnerable to these pressures are

1. Administration of tests by computers,
2. Computer-assisted test assembly,
3. The generation of items by computer.

Administration of Tests

The administration of tests by computers is no longer just a possibility, it is a reality. Moreover, it stands as one of the proudest achievements of psychometrics because the theory that would make adaptive testing a reality, Item Response Theory (IRT; Lord, 1980), was developed before computers were widely available. Had this theory not been developed it is likely that in the current technological revolution computers would have been applied to testing in a shallow manner. That is, computers probably would have used as automated answer sheets rather than as a means of delivering new kinds of tests or more efficient tests.

By the early 1970s computer technology had reached the point where it was possible simultaneously to test several examinees more or less economically. The pioneering efforts of Weiss (1974) capitalized on this event and on the availability of IRT to begin an extensive research program on the psychometric and practical issues of adaptive testing.

In an adaptive test, the computer's job is not merely to present the item and score it but also to determine which item should be administered next, given the student's current level of performance. Although adaptive tests usually use multiple-choice items and thus give the impression that a paper-and-pencil test has been transferred to a computer, in reality different examinees are responding to different tests assembled by the computer for each examinee so that the resulting score may be most precise for an individual test taker.

It is tempting to say that adaptive testing became possible as a result of coupling computers and IRT. The fact is that Binet was doing pretty much the same thing at the turn of this century. Of course, then it was the psychometrician, not the computer, that was selecting and scoring the items. Adaptive testing is thus an efficient implementation of a long-standing idea. Nevertheless, it is still a significant achievement, especially considering what would have happened in the absence of IRT--namely, the blind transfer of items to a computer screen. That achievement is about to become a practical reality. The military and private testing organizations have both been seriously contemplating the practical implementation of adaptive testing systems. In some cases concrete steps have already been taken toward their implementation. Although it is too early to tell what success these initial efforts will encounter, computers are becoming so pervasive that not to give a test by computer may soon appear archaic. Chances are that there will thus be more computer administration of tests, although not necessarily because they are better psychometrically. It will therefore be up to the test designer to make the best possible use of the available technology.

While adaptive testing has been moving forward, technology, psychometrics and substantive theory have not remained static, and the integration of these three opens up additional opportunities. For example, most adaptive testing research has been limited to verbal items. This was so because until recently it was too expensive to display symbols and graphics on a CRT (cathode ray tube). That has changed and in principle test material can even be presented in the form of television images by means of videodisc players. A videodisc permits access to up to 54,000 television frames and, for example, language skills could be tested in very realistic contexts by presenting items as audiovisual sequences. On the psychometric front, models that go beyond the classification of responses into "correct" and "incorrect" have been formulated (e.g., Andersen, 1977; Bock, 1972; Fischer, 1973; Samejima, 1969; 1972; Embretson, Chapter 7, this volume; Scheiblechner, Chapter 8, this volume; Andrich, Chapter 9, this volume) but await tests that make use of their capacity. Finally, on the theoretical side, experimental psychologists have taken seriously Cronbach's exhortation (Cronbach, 1957) to unite experimental and differential psychology. As a result, there have been serious attempts since the 1960s to understand test performance in the light of substantive, not just quantitative, theories (e.g., Carroll, 1976; Embretson, 1983; Lansman, Donaldson, Hunt, & Yantis, 1982). In short, the materials are there not only to improve current practice but also to chart new courses.

<u>Computers and Test Assembly</u>

In the 1960s, one would have predicted that the computer's first inroad into test design would be in assisting with the test development

process rather than in administering tests. As just shown, however, test administration by computer is becoming a reality. By contrast, the possibilities of using computers for test assembly and test creation have hardly been exploited. Before speculating on how test assembly and item generation can benefit from the integration of psychometrics, technology, and psychological theory, I first review the state of the art.

A key problem in test development is maintaining a large item pool from which items may be drawn, according to some set of specifications, to assemble the final form. For the most part, item pools are kept in filing cabinets. When the time comes, however, to assemble another form it might be wise to sweep the floor, because often the test assembler spreads the cards on the floor to select (through an as yet unpublished procedure) a set of items. Typically, the items in the pool have been pretested, a requirement imposed by the actuarial nature of test development. Of course, that is not the end of the process. Once a tentative set of items has been chosen it goes through numerous revision stages in which some items are deleted and still others added. The criteria for reviewing items include the following:

1. Distribution of item statistics such as difficulty and discrimination,
2. distribution of distractors,
3. lexical overlap,
4. conceptual overlap,
5. content classifications,
6. ethnic, racial, and gender bias.

Some of these criteria involve only surface characteristics of the items. Were it not for the fact that items are usually pretested, the test assembler would often have an erroneous idea of the difficulty and discrimination of the item (e.g., Bejar, 1983a). It is in this sense that current test design is an actuarial science. Precisely because tests are assembled on the basis of surface characteristics, the process is amenable to computerization. Such computerization will take place if for no other reason than that it increases productivity.

For example, computers can, to the extent that the item pool permits, assemble a form to meet the requirements just enumerated while simultaneously attempting to meet some psychometric criterion, such as the distribution of item difficulty and discrimination. The ideal system would be flexible enough to accommodate the styles of different test designers, and it would also be interactive. For example, the system should present the test designer with the option of either letting the computer suggest a form or allowing the test designer to assemble a form gradually. In either case the system should be interactive in the sense of allowing the test designer to ascertain how well the design goals have been met as often as the test designer desires. Naturally, the system should be powerful enough to access sizable item pools instantly, regardless of their graphic complexity.

Components of some of these ideas are being contemplated or in some cases have been implemented (e.g., Yen, 1983), but clearly there is room for improvement. For example, while the computer is in the process of selecting a set of items it may easily produce a report on the

availability of different item types for the test designer--who in turn, could take the necessary steps to replenish the item pool following the suggestions of the computer. It is at this point, however, that the actuarial nature of current test design makes itself obvious. If, for example, the computer reports that easy items of a certain category are running out the test designer can, at best, make the arrangements to pretest another batch of items and hope that among them there will be a large number of easy items.

A system to implement these ideas, to my knowledge, has neith been developed nor is it under serious consideration. It is, howe question of time before the economics of the present labor-intensive approach becomes unbearable. Because substantial planning is required to develop such a system, it would be desirable to begin now before the need becomes urgent.

Using the Computer to Generate Items

Computers can be useful for test design because they can advise the test designer about the characteristics of unpretested items and, ultimately, to generate items according to a prescription. These activities would, of course, be much more difficult to achieve; moreover, it would make the system described earlier unnecessary because in generating items the computer would make sure that they meet the required specifications. That is, rather than maintaining large item pools, as is now done, a point may be reached where submitting a prescription for a test to a computer that would produce a test meeting all the content and psychometric specifications would be feasible. Are

we anywhere near the point where such feats are possible? A brief review of the state of the art is very much in order at this juncture.

The essence of the item generation process as it is currently practiced was described by Wesman (1971):

> Item writing is essentially creative--it is an art. Just as there can be no set of formulas for producing a good story or a good painting, so there can be no set of rules that guarantees the production of good test items. Principles can be established and suggestions offered, but it is the writer's judgement in the application--and occasional disregard--of these principles and suggestions that determines whether good items or mediocre ones are produced. Each item, as it is being written, presents new problems and new opportunities. Thus item writing requires an uncommon combination of special abilities and is mastered only through extensive and critically supervised practice. (p. 81)

Chances are good that the state of affairs described by Wesman will prevail in the immediate future. However, some efforts (e.g., Roid & Haladyna, 1982) are under way to make item writing more a science than an art. However, the foundation of many of the procedures outlined in the Roid and Haladyna work rest on a behaviorist foundation, which may make them incompatible with the cognitive turn that psychology and psychometrics have taken. For example, one item generation technique that has evolved is the <u>item form</u> (Hively, 1974). Hively defined an item form as a list of rules for generating a set of items. An item in turn is defined as a "set of instructions telling how to evoke, detect

and score a specific bit of human performance. It must include the directions for (1) presenting the stimuli, (2) recording the response, and (3) deciding whether or not the response is appropriate" (Hively, 1974, p. 8).

From a psychometric and technological standpoint, item forms are attractive. They are congenial test development procedures for psychometric models relying on the assumption that the items in a test are a random sample from some universe of item. Generalizability theory (Brennan, 1983; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) is the most prominent model based on that assumption. From a technological point of view, item forms are also attractive because they permit a computer to generate items. That is, the item form can be viewed as a program that in principle can enumerate all the items that belong to the universe. By the random choice of items from this universe a test can be formed that satisfies the random sampling assumption. Although item forms and generalizability theory are very compatible, the psychometrics of behavioristically oriented test design has often taken the form of very specific models (e.g., Harris, Pastorok, & Wilcox, 1977) rather than the broader foundation provided by generalizability theory.

In short, the closest that has been come to using computers for item generation is through the notion of an item form from which a universe of items can be generated. In my estimation, that approach to item generation is too specialized. In practice, items differ with respect to a number of characteristics, and a useful generation scheme must have control over those characteristics. For example, a useful generation scheme should be able to generate easy items or hard items at

will. I suspect that to build such systems it is first necessary to have an idea of what makes an item easy c hard. Some insights on beginning to do this can be found in efforts concerned with the development of computer programs that take tests (see, e.g., Evans, 1968; Green, 1964; Simon & Siklossy, 1972).

Cognitive Science and Psychometrics

A quick review of the history of psychology (e.g., Boring, 1950) shows that throughout the history there has been a tension between the study of consciousness and the study of behavior. As Boring put it, "in its simplest terms the basic problem about the data of psychology is this: Does psychology deal with the data of consciousness or data of behavior or both?" (Boring, 1950, p. 620). These tensions between opposing views often manifest themselves in psychology, as well as in other sciences, in the form of dichotomies (Newell, 1983). Within psychology behaviorism once dominated the field. The pendulum has now swung and mentalism, in the form of cognitive psychology, now has the upper hand. It seems that psychometrics has swung along with the rest of psychology, as evidenced by the vigor of efforts to cognitivize psychometrics. Some of these efforts are represented in this volume. (The reader is also referred to Embretson, 1983, for an approach that encompasses not only test design, which she calls construct representation, but also an accounting of the relationship among scores from several tests, which she calls nomothetic span.)

It is not necessary to feel sorry for the behaviorist. When behaviorism was champion, psychometricians of that persuasion had their

day, as demonstrated by the following excerpt from Osburn (1968) regarding test design.

> Few measurement specialists would quarrel with the premise that the fundamental objective of achievement testing is generalization. Yet the fact is that current procedures for the construction of achievement tests do not provide an unambiguous basis for generalization to a well defined universe of content. At worst, achievement tests consist of arbitrary collections of items thrown together in a haphazard manner. At best, such tests consist of items judged by subject matter experts to be relevant to and representative of some incompletely defined universe of content. In neither case can it be said that there is an unambiguous basis for generalization. This is because the method of generating items and the criteria for the inclusion of items in the test cannot be stated in operational terms.
>
> The time-honored way out of this dilemma has been to resort to statistical and mathematical strategies in an attempt to generalize beyond the arbitrary collection of items in the test. By far the most popular of these strategies has been to invoke the concept of a latent variable--an underlying continuum which represents a hypothetical dimension of skill. (p. 95)

The notion of criterion-referenced tests was popularized by Glaser and Nitko (1971) shortly thereafter, and for over a decade criterion-referenced tests enjoyed the endorsement of many psychometricians and

clearly had an impact on test design (see Shoemaker, 1975). It is perhaps no coincidence that critics, once behaviorism ceased to be a major influence in psychology, began finding all sorts of problems in criterion-referenced tests. For example, Johnson and Pearson (1975) criticized criterion-referenced reading tests as being linguistically naive. They argued that by focusing exclusively on observable interpretations the usefulness of measuring instruments is diminished. Moreover, advocates of criterion-referenced measurement (e.g., Hambleton, Swaminathan, Algina, & Coulson, 1978; Nitko, 1980) have begun to accept construct validation as playing a useful role in the validation of criterion-referenced tests. This of course implies their acceptance of the legitimacy of using nonobservable constructs in test interpretation. Indeed, there is no reason why an emphasis on behavior and cognition cannot coexist in both an instructional and a psychometric sense (Greeno, 1978).

The more recent emphasis on cognitive psychology has at least two implications for psychometrics. One is the possibility of understanding test performance in terms of cognitive constructs (e.g., Sternberg, 1981). The other possibility is the exploitation of cognitive theory for the improvement and design of both currrent and fundamentally new tests. In the next section I discuss both possibilities.

Validation of Test Performance

The most likely immediate influence of cognitive science on psychometrics is as a source of constructs to validate test scores. Messick (1975) has eloquently argued for the necessity of construct

validation, and the case need not be repeated here. It is sufficient to say that the availability of cognitive or information-processing constructs and the revival of construct validation have important implications for test design.

The validation of both aptitude and achievement tests has relied very little on cognitive constructs. In the recent past validation of achievement tests was strongly influenced by content considerations. This was in line with the behavioristic orientation of criterion-referenced testing that has dominated much of the thinking in the field. Similarly, the validation of aptitude tests, from the Scholastic Aptitude Test (SAT) to the Armed Services Vocational and Aptitude Battery (ASVAB) has relied almost exclusively on predictive validity, and this paradigm is responsible for the psychometric nature of procedures for improving validity. The alternative view is that understanding the nature of the relationship, as opposed to just its magnitude, puts test developers in a better position to increase validity. However, validation based on cognitive constructs, and for that matter tests developed from scratch based on cognitive theory, need not necessarily yield higher predictive validities. It is known from psychometric theory that the magnitude of correlation between a test and a criterion is determined by the proportion of variance in common between the two. Clearly, the test designer has control over the composition of the test but not over the composition of the criterion. Hunt (1983) anticipated this when he noted the following:

> The cognitive science view may lead to the development of new tests that are more firmly linked to a theory of cognition than are present tests. Such tests are yet to be written.

> There is no compelling reason to believe that new tests will be better predictors of those criteria that are predicted by today's tests. After all the present tests are the results of an extensive search for instruments that meet the pragmatic criterion of prediction. Theoretically based tests may expand the range of cognitive functions that are evaluated and certainly should make better contact with our theories of cognition. Theoretical interpretation, alone, is not a sufficient reason for using a test. *A test that is used to make social decisions must meet traditional psychometric criteria for reliability and validity* [italics added]. No small effort will be required to construct tests that meet both theoretical and pragmatic standards. The effort is justified, for our methods of assessing cognition ought to flow from our theories about the process of thinking. (p. 146)

Moreover, from a social perspective, validation solely in terms of predictive validity is inadequate. A predictive validation strategy may have been appropriate when the primary object of testing was the identification of high-scoring individuals, but society's concern with equality requires a focus on low-scoring individuals also. As noted by the Committee on Ability Testing of the National Research Council:

> The relationship between problem solving on tests and everyday performance has taken on new relevance to public policy, as attention has come to focus. . . not on those selected, as was the case when tests were perceived primarily as identifying

> excellence, but on those not selected. This shift in focus has brought new prominence to the question of what is being measured by a given test or item type and has pointed up insufficiencies from a public perspective in validation strategies based solely on the demonstration of external statistical relationships. (Wigdor & Warner, 1982, p. 215)

This quotation and much of the litigation involving tests suggest that in the years ahead test designers will have to be more sensitive to the ethical implications of testing instruments. That is, test designers will have to take into account not just the psychometric and substantive base of tests but their consequences as well. Messick (1980) has suggested that the consequences of testing should be a component of the validation process rather than an afterthought. Just as construct validation consists of collecting evidence from many substantive perspectives, the procedure for incorporating consequences into the validation process consists of collecting information on the implications of using a test in a particular situation. However, such listing of implications cannot be fruitfully done in a psychometric vacuum:

> Appraising the possible consequences of test use is not a trivial process under any circumstances, but it is virtually impossible in the absence of construct validity information about the meaning and nature of test scores. Just as the construct network of nomological implications provided a rational basis for hypothesizing potential relationships to criteria, so it also provides a rational

basis for hypothesizing potential outcomes and for anticipating possible side effects. (Messick, 1980, p. 15)

An Illustration

An example of construct validation in the context of an adaptive test is provided by Bejar and Weiss (1978). They postulated a nomological net to account for achievement in a college biology course and proceeded to test its feasibility with a structural equation model (see Bentler, 1978, for a discussion of construct validation by means of structural equation models). The net is seen in Fig. 1. The rectangles represent the constructs postulated to account for the relationships among the six observable variables. The coefficients next to the arrows are those that need to be estimated. The direction of the arrow indicates that the variable at the head of the arrow is regressed on the variable at the other end of the arrow. Bejar and Weiss concluded that the postulated model indeed fitted the data and that although there were no major differences between the validity of paper-and-pencil and adaptive versions of the test, the adaptive test required 25% fewer items. At the time, such a reduction in the number of items, compared to the cost of an adaptive test, may not have been cost-effective. By the 1980s, of course, the hardware cost per terminal could have easily been less than $1,000, and the economics of adaptive testing may thus appear more attractive.

(Insert Figure 1 about here)

The net postulated by Bejar and Weiss was dictated more by the availability of scores than by an information-processing model of achievement. If measures inspired by cognitive science had been

available they could easily have been used. Rose (1980), for example, developed a battery of tasks that are indicators of information processes. The use of that battery in the validation of the adaptive biology achievement test would have been consistent both with what has been called a cognitive correlates and cognitive components approach to cognitive psychometrics (see Pellegrino & Glaser, 1979; Sternberg, 1981). In the cognitive correlates approach, the goal is to test subjects on several low-level tasks that are believed to be indicative of the subjects' efficiency in processing information. An example of a low-level task is matching whether two letters, such as Cc, constitute a physical match or, as in this case, a name match. Because the tasks are easy, response latency, rather than correctness, is the outcome of interest on such tasks. In a cognitive components approach, the aim is to postulate a model of information processing and to test by obtaining data on the performance of subjects on testlike tasks. The outcomes from either approach can be used as part of a construct validation study designed to gain further understanding of the performance of students in a test.

Although it is beyond the scope of this chapter to consider achievement testing in detail (see Bejar, 1983b), it should be noted that performance on an achievement test depends on both processing components and the storing of information. The cognitive components and correlates approach emphasizes the processing part but not the storage part, that is, the schema for representing information. Other researchers (e.g., Burton, 1982) have emphasized the storage part by elaborating constructs about how the students represent knowledge.

The Bejar and Weiss (1978) study, in addition to illustrating what Messick (1980) has called evidential component of validation, also

illustrated the consequential aspect of validation. Bejar and Weiss found evidence in their data of a medium effect. That is, it seemed that the medium of administration, whether it was paper and pencil or computerization, influenced the scores to some extent. If this medium effect can be replicated, the possible ethical consequences will be something for future test designers to worry about. For example, students from less-affluent homes are less likely to have been exposed to keyboards and CRTs and thus may obtain lower scores. Because these students are also likely to have been exposed to a less-adequate educational environment, it would add insult to injury to test those students by computer without first ensuring that they are at ease with the computer as a medium for test delivery. The work of Snow and Peterson (Chapter 5 in this volume) has obvious implications for research on the detection of such problems.

Towards Scientifically Based Test Design

In the previous section I discussed construct validation as a means to a better understanding of test scores. Although, no doubt, such information could be useful to a test developer, he or she may be at a loss on how to incorporate that information in the creation of new items or of entirely new tests. In this section I argue that from a test design perspective it is necessary to shift the focus of attention from the examinee to the item. That is, just as construct validation of test scores entails research to understand differences among examinees, construct validation applied to test design entails research to understand differences among items. More concretely, it is necessary to

account for the differences among items with respect to their characteristics, especially difficulty. I suggest that cognitive science is an important source of ideas for accomplishing that goal. This integration of psychometric models and cognitive science, as reflected in the work of Embretson (1983) and Fischer (1973), is important not only for advancing the scientific status of psychometric instruments but also for creatively incorporating technological advances into the testing process. For example, if test developers are able to account for differences among items they may have captured the knowledge necessary to synthesize items of known characteristics (see Egan, 1979). They may, in short, be able to write a computer program capable of composing an item with known psychometric characteristics. The chapters by Sternberg and McNamara (Chapter 2), Pellegrino, Mumaw, and Cantoni (Chapter 3), and Butterfield, Nielsen, Tangen, and Richardson (Chapter 4) in this volume provide a basis for research toward that goal.

I would be the first to agree that synthesizing items is not likely to be easy and that sustained research is required before practical results will be available. Nevertheless, adopting that effort as a goal puts test developers in the enviable position of simultaneously pursuing scientific and economic goals. That is, the ability to synthesize items is likely to improve the productivity of the test designer in much the same way that computers have altered the productivity of, for example, graphics designers in various industries. To reach that point, however, they will have to do considerable work to establish and validate a theory that explains the characteristics of items.

It is beyond the scope of this chapter to outline a detailed research program that will in the end allow the synthesis of items.

However, a natural starting point is to account for the variability among existing items. (See Carroll, 1979 for an attempt to do so.) Unfortunately, this task is made difficult by the fact that most existing tests are of the multiple-choice variety. No doubt with such items the context in which the correct alternative occurs partially determines the psychometric characteristics of the item. This, unfortunately, makes the task more difficult than it ought to be because the multiple-choice item was invented to facilitate group testing, and thus its usefulness will presumably diminish as computers are used more and more in the administration of individualized tests. In the meantime, however, test designers must be ready to deal with the complications introduced by multiple-choice items.

Psycholinguistic theory is a rich source of hypotheses for the study of verbal tests such as reading comprehension tests and writing ability tests. Psychologists have devoted considerable attention to sentence comprehension (e.g., Kintsch, 1977). One early theory was postulated by Miller (Miller & McKeon, 1964) and is known as the Derivational Complexity Theory. According to this theory the comprehensibility of a sentence is determined by the syntactic complexity of the sentence. Complexity was measured as the number of transformations required to go from the deep structure to the surface structure of a sentence. Although this particular theory is not now well supported, it seems reasonable to suggest that if comprehensibility of a sentence is affected by some measure of syntactic and semantic complexity then psychometric difficulty of an item based on that sentence will to some extent also depend on the syntactic and semantic complexity of the sentence.

A test with items based on sentences is the Test of Standard Written English (TSWE), sponsored by the College Board and produced by the Educational Testing Service. One of the two item types in the test consists of a sentence that may or may not contain a grammatical error. The examinee's task is to determine whether the sentence as it stands contains an error; if it does, the examinee must select from several alternatives to correct the sentence. One way to apply these ideas to the TSWE is to obtain several measures of linguistic complexity on each item and study the relationship of those measures to psychometric difficulty. If a stable relationship is found, then, in principle, the resulting model may be used to predict the difficulties of new items and even to modify items so they will be easier or harder.

Although the preceding remarks are speculative, some research along these lines already exists. For example, the Degrees of Reading Power (DRP) sponsored by the College Board, is a reading comprehension cloze test. Unlike the usual cloze test, the DRP is a multiple-choice test; that is, the examinee is provided several choices for filling in the deleted word. The difficulty of those items can apparently be predicted on the basis of the readability index of the passage. Similarly, Swinton (personal communication) has experimented with verbal analogy items by forming different versions of the item in order to alter their difficulty.

The idea of synthesizing items of known characteristics has been implemented by at least one research team (Burton, 1982). They were concerned with the design of diagnostic tests of subtraction. Their

goal was to infer what misconceptions may account for a student's error in arithmetic. To speed that process up it is necessary to synthesize items "on the fly" that are most informative with respect to the current set of hypothesized misconceptions. That is, a computer creates the items as they are needed rather than retrieving them from a pool.

One area that seems ready for the integration of cognitive theory and psychometric models is spatial ability. Spatial ability has been a subject of intense investigation. A well-established finding is that the response latency to problems that require mental manipulation is a function of the physical characteristics of the test stimuli. For example, the time it takes to determine whether two geometric figures are the same is a linear function of their angular disparity (see Cooper, 1980). This finding suggests that the psychometric difficulty of spatial items could be predicted from an analysis of their physical characteristics. A project investigating this possibility is under way at Educational Testing Service under the sponsorship of the Office of Naval Research.

Concluding Comments

In this chapter I have attempted to enumerate some of the ways in which the integration of technology, cognitive science, and psychometric theory can benefit test design. The state of the art is most advanced with respect to the administration of tests, with the notion of adaptive tests rapidly approaching operational implementation. As I have suggested, adaptive testing is a significant step forward. However, from a user's point of view, an adaptive test is just a multiple-choice

test administered by computer because the improvements in efficiency and even the test's psychological advantages are not obvious to the naked eye.

I have argued that to move the state of the art forward it will be necessary to pay closer attention to the psychological foundation of tests. This effort calls, on one hand, for the construct validation of tests from both an evidential and consequential perspective. On the other hand, I have also argued that to improve the scientific basis of test design it is necessary to focus attention not only on variability among examinees but also on variability among items. In particular, a better understanding of why items behave the way they do is needed. From a practical perspective the payoff for doing so will be the possibility of ultimately being able to synthesize items of known psychometric characteristics.

References

Andersen, E. B. (1977). The logistic model for m-answer categories. In W. F. Kempf & B. H. Repp (Eds.), Mathematical models for social psychology. New York: Wiley.

Bejar I. I. (1983a). Subject-matter expert prediction of item difficulty. Applied Psychological Measurement, 7, 303-310.

Bejar, I. I. (1983b). Achievement testing: Recent developments. Beverly Hills: Sage Publications.

Bejar, I. I., & Weiss, D. J. (1978). A construct validation of adaptive achievement testing (Research Rep. 78-4). Minneapolis: University of Minnesota, Department of Psychology.

Bentler, P. M. (1978). The interdependence of theory methodology and empirical data: Causal modeling as an approach to construct validation. In D. B. Kandel (Ed.), Longitudinal research on drug use. New York: Halstead Press.

Bock R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 38, 437-458.

Boring, E. G. (1950). A history of experimental psychology. New York: Appleton-Century-Crofts.

Brennan, R. L. (1983). Elements of generalizability theory. Iowa City: ACT Publications.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems. New York: Academic Press.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new structure of intellect. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

Carroll, J. B. (1980). Measurement and ability constructs. In Construct validity in psychological measurement. Proceedings of a Colloquium on theory and application in education and employment. Princeton, NJ: Educational Testing Service.

Cooper, L. A. (1980). Spatial information processing: Strategies for research. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), Aptitude, learning and instruction (Vol. 1). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Egan, D. F. (1979). Testing based on understanding: Implication from studies of spatial ability. Intelligence, 3, 1-15.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. Psychological Bulletin, 93, 179-197.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 36, 359-374.

Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), _Educational measurement_. Washington, DC: American Council on Education.

Green, B. F. (1964). Intelligence and computer simulation. _Transactions of the New York Academy of Sciences_, Ser II, _27_, 55-63.

Greeno J. G. (1978). Book review of Human characteristics and school learning, by B. Bloom. _Journal of Educational Measurement_, _15_, 67-76.

Hambleton R. K., Swaminathan, H., Algina, J., & Coulson, D. B. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. _Review of Educational Research_, _48_, 1-47.

Harris C. W., Pastorok, A., & Wilcox, R. R. (1977). _Achievement testing: Item methods of study_. Los Angeles: Center for the Study of Evaluation, University of California.

Hively, W. (1974). Introduction to domain referenced testing. _Educational Technology_, _14_, 5-9.

Hunt, E. (1983, January). On the nature of intelligence. _Science_, _219_ (No. 4581), 141-146.

Johnson, D., & Pearson P. D. (1975, May). Skill management systems: A critique. _The Reading Teacher_, 757-764.

Kintsch, W. (1977). _Memory and cognition_. New York: Wiley.

Lansman, M., Donaldson, G., Hunt, E., & Yantis, S. (1982). Ability factors and cognitive processes. _Intelligence_, _6_, 347-386.

Lord, F. M. (1980). _Application of item response theory to practical testing problems_. Hillsdale, NJ: Erlbaum.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Miller, G. A., & McKeon, K. O. (1964). A chronometric study of some relations between sentences. Quarterly Journal of Experimental Psychology, 16, 297-308.

Newell A. (1983). Intellectual issues in the history of artificial intelligence. In F. Machlup & U. Mansfield (Eds.), The study of information: Interdisciplinary messages. New York: Wiley.

Nitko, A. J. (1980). Distinguishing the many varieties of criterion-referenced tests. Review of Educational Research, 50, 461-486.

Osburn, H. G. (1968). Item sampling for achievement testing. Educational and Psychological Measurement, 28, 95-104.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components on the analysis of individual differences. Intelligence, 3, 187-214.

Roid, G., & Haladyna, T. (1982). A technology for test item writing. New York: Academic Press.

Rose, A. M. (1980). Information processing abilities. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), Aptitude, learning and instruction (Vol. 1). Hillsdale, NJ: Erlbaum.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded responses. Psychometrika, Monograph No. 17.

Samejima, F. (1972). A general model for free response data. Psychometrika, Monograph No. 18.

Shoemaker, D. M. (1975). Towards a framework in assessment testing. Review of Educational Research, 48, 127-141.

Simon, H. A., & Siklossy, L. (Eds.). (1972). Representation and meaning: Experiments in information processing systems. Englewood Cliffs, NJ: Prentice-Hall.

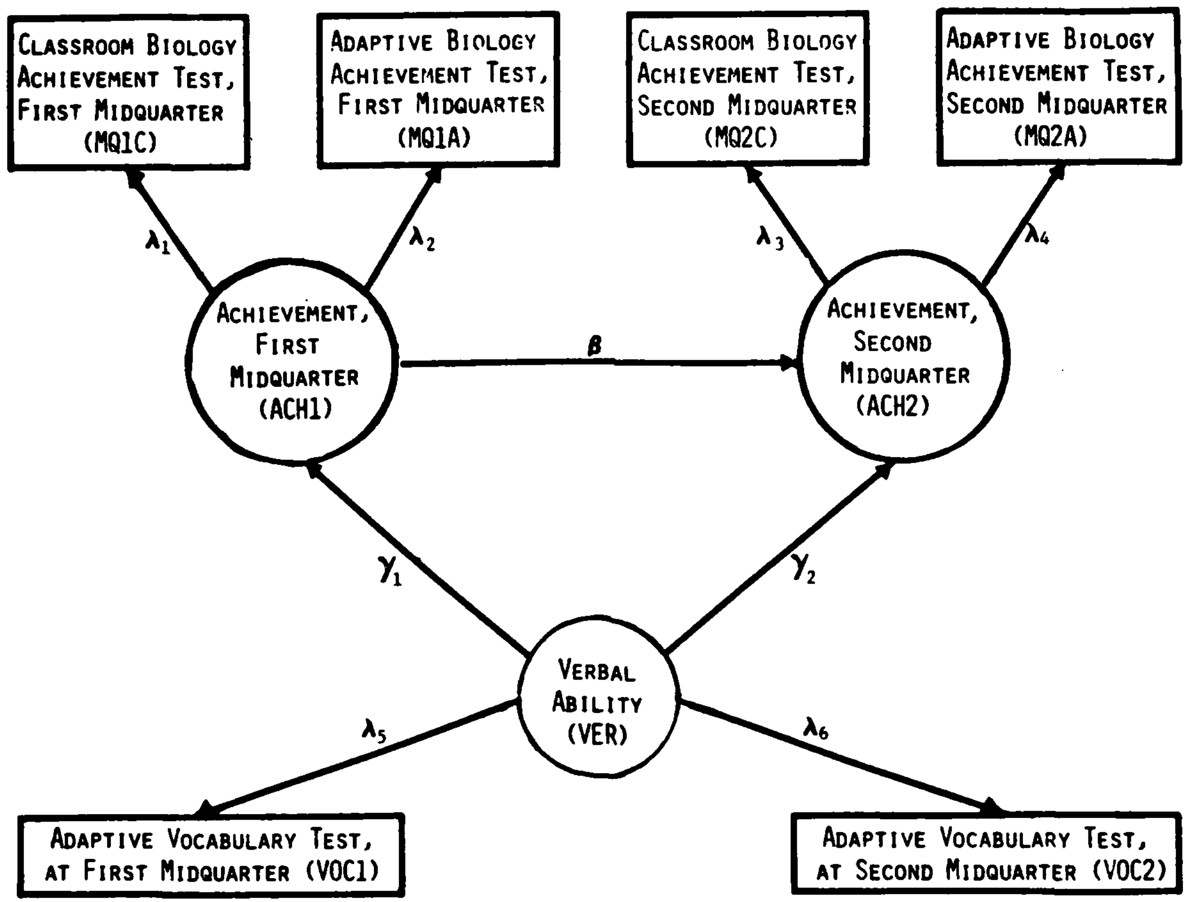
Sternberg, R. J. (1981). Testing and cognitive psychology. American Psychologist, 36, 1181-1189.

Weiss, D. J. (1974). Strategies of adaptive ability measurement (Research Rep. 74-5). Minneapolis: University of Minnesota, Department of Psychology.

Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), Educational measurement. Washington, DC: American Council on Education.

Wigdor, A. K., & Warner, W. R. (Eds.) (1982). Ability testing: Uses, consequences, and controversies (Part II). Washington, DC: National Academy Press.

Yen, W. (1983). Use of the three-parameter model in the development of a standarized achievement test. In R. K. Hambleton (Ed.), Applications of item response theory. Vancouver: Educational Research Institute of British Columbia.



Classroom Biology Achievement Test, First Midquarter (MQ1C)
Adaptive Biology Achievement Test, First Midquarter (MQ1A)
Classroom Biology Achievement Test, Second Midquarter (MQ2C)
Adaptive Biology Achievement Test, Second Midquarter (MQ2A)
λ1
λ2
λ3
λ4
Achievement, First Midquarter (ACH1)
β
Achievement, Second Midquarter (ACH2)
γ1
γ2
Verbal Ability (VER)
λ5
λ6
Adaptive Vocabulary Test, at First Midquarter (VOC1)
Adaptive Vocabulary Test, at Second Midquarter (VOC2)

Navy

1 CDR Robert J. Biersner
Naval Medical R&D Command
National Naval Medical Center
Bethesda, MD 20814

1 Dr. Nick Bond
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

1 Lt. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32508

1 Dr. Robert Carroll
NAVOP 115
Washington , DC 20370

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. Cathy Fernandes
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508

1 Dr. Kneale Marshall
Chairman, Operations Research Dept.
Naval Post Graduate School
Monterey, CA 93940

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

1 Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

Navy

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode, Director
Department N-7
Naval Training Equipment Center
Orlando, FL 32813

1 Dr. Richard Snow
Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 Jerry Lehnus
CAT Project Office
HQ Marine Corps
Washington , DC 20380

1 Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. James Baker
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Blvd.
Alexandria , VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

1 Commander, U.S. Army Research Institute
for the Behavioral & Social Sciences
ATTN: PERI-BR (Dr. Judith Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joyce Shields
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Army

1 Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Air Force

1 Technical Documents Center
Air Force Human Resources Laboratory
WPAFB, OH 45433

1 U.S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. Patrick Kyllonen
AFHRL/MOE

Brooks AFB, TX 78235

1 Dr. Roger Pennell
Air Force Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Dr. Joseph Yasatuke
AFHRL/LRT
Lowry AFB, CO 80230

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defens
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. W. Steve Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209

1 Dr. Robert A. Wisher
OUSDRE (ELS)
The Pentagon, Room 3D129
Washington, DC 20301

Civilian Agencies

1 Dr. Patricia A. Butler
NIE-BRN Bldg. Stop # 7
1200 19th St., NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Office of Scientific and Engineering
Personnel and Education
National Science Foundation
Washington, DC 20550

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 326

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 Psychological Research Unit
NBH-3-44 Attn
Northbourne House
Turner ACT 2601
AUSTRALIA

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Ina Bilodeau
Department of Psychology
Tulane University
New Orleans, LA 70118

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. Werner Birke
Personalstammamt der Bundeswehr
D-5000 Koeln 90
WEST GERMANY

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52243

1 Dr. Glenn Bryan
6208 Poe Road
Bethesda, MD 20817

Private Sector

1 Bundministerium der Verteidigung
-Referat P II 4-
Psychological Service
Postfach 1328
D-5300 Bonn 1
F. R. of Germany

1 Dr. Ernest R. Cadotte
307 Stokely
University of Tennessee
Knoxville, TN 37916

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

1 Dr. Walter Cunningham
University of Miami
Department of Psychology
Gainesville, FL 32611

1 Dr. Dattpradad Divgi
Syracuse University
Department of Psychology
Syracuse, NE 33210

Private Sector

1 Dr. Emmanuel Donchin
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Hei-Ki Dong
Ball Foundation
Room 314, Building B
800 Roosevelt Road
Glen Ellyn, IL 60137

1 Dr. Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel St.
Champaign, IL 61820

1 Dr. Susan Embertson
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Dexter Fletcher
University of Oregon
Department of Computer Science
Eugene, OR 97403

Private Sector

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01002

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

1 Glenda Greenwald, Ed.
Human Intelligence Newsletter
P. O. Box 1163
Birmingham, MI 48012

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
603 East Daniel Street
Champaign, IL 61820

Private Sector

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Advanced Statistical Technologies
Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

1 Dr. Marcel Just
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Professor John A. Keats
Department of Psychology
The University of Newcastle
N.S.W. 2308
AUSTRALIA

1 CDR Robert S. Kennedy
Canyon Research Group
1040 Woodcock Road
Suite 227
Orlando, FL 32803

1 Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

1 Dr. Stephen Kosslyn
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138

Private Sector

1 Dr. Marcy Lansman
The L. L. Thurstone Psychometric Laboratory
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Mr. Phillip Livingston
Systems and Applied Sciences Corporatio
6811 Kenilworth Avenue
Riverdale, MD 20840

1 Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

1 Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Private Sector

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Don Lyon
P. O. Box 44
Higley , AZ 85236

1 Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

1 Dr. Scott Maxwell
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Mr. Robert McKinley
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

1

Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Robert Mislevy
711 Illinois Street
Geneva, IL 60134

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Private Sector

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara , CA 93106

1 Dr. Steven E. Poltrock
Bell Laboratories 2D-444
600 Mountain Ave.
Murray Hill, NJ 07974

1 Dr. Mike Posner
Department of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

1 Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P. O. Box 688
Richardson, TX 75080

Private Sector

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Frank L. Schmidt
Department of Psychology
Bldg. GG
George Washington University
Washington, DC 20052

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

1 Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

1 Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Private Sector

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Kathryn T. Spoehr
Psychology Department
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Martha Stocking
Educational Testing Service
Princeton, NJ 08541

1 Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Private Sector

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. V. R. R. Uppuluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

1 Dr. David Vale
Assessment Systems Corporation
2233 University Avenue
Suite 310
St. Paul, MN 55114

1 Dr. Kurt Van Lehn
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin--Milwaukee
Milwaukee, WI 53201

1 Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Private Sector

1 Dr. Donald O. Weitzman
Mitre Corporation
1820 Dolley Madison Blvd
McLean, VA 22102

1 William B. Whitten
Bell Laboratories
2D-610
Holmdel, NJ 07733

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90007

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

END

FILMED

8